8-10-2010

Gene cooption and convergent evolution of oxygen transport hemoglobins in jawed and jawless vertebrates

Federico G. Hoffmann
University of Nebraska - Lincoln, fhoffmann2@unl.edu

Juan C. Opazo
University of Nebraska - Lincoln, jopazo2@unl.edu

Jay F. Storz
University of Nebraska - Lincoln, jstorz2@unl.edu

Gene cooption and convergent evolution of oxygen transport hemoglobins in jawed and jawless vertebrates

Federico G. Hoffmann[a], Juan C. Opazo[b], and Jay F. Storz[a,1]

[a]School of Biological Sciences, University of Nebraska, Lincoln, NE 68588; and [b]Instituto de Ecología y Evolución, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile

Edited* by Morris Goodman, Wayne State University, School of Medicine, Detroit, MI, and approved July 6, 2010 (received for review May 14, 2010)

Natural selection often promotes evolutionary innovation by coopting preexisting genes for new functions, and this process may be greatly facilitated by gene duplication. Here we report an example of cooptive convergence where paralogous members of the globin gene superfamily independently evolved a specialized O_2 transport function in the two deepest branches of the vertebrate family tree. Specifically, phylogenetic evidence demonstrates that erythroid-specific O_2 transport hemoglobins evolved independently from different ancestral precursor proteins in jawed vertebrates (gnathostomes) and jawless fish (cyclostomes, represented by lamprey and hagfish). A comprehensive phylogenetic analysis of the vertebrate globin gene superfamily revealed that the erythroid hemoglobins of cyclostomes are orthologous to the cytoglobin protein of gnathostome vertebrates, a hexacoordinate globin that has no O_2 transport function and that is predominantly expressed in fibroblasts and related cell types. The phylogeny reconstruction also revealed that vertebrate-specific globins are grouped into four main clades: (*i*) cyclostome hemoglobin + cytoglobin, (*ii*) myoglobin + globin E, (*iii*) globin Y, and (*iv*) the α- and β-chain hemoglobins of gnathostomes. In the hemoglobins of gnathostomes and cyclostomes, multisubunit quaternary structures provide the basis for cooperative O_2 binding and allosteric regulation by coupling the effects of ligand binding at individual subunits with interactions between subunits. However, differences in numerous structural details belie their independent origins. This example of convergent evolution of protein function provides an impressive demonstration of the ability of natural selection to cobble together complex design solutions by tinkering with different variations of the same basic protein scaffold.

cytoglobin | gene family evolution | globin | hagfish | lamprey

Natural selection often promotes evolutionary innovation by coopting preexisting genes for new functions. Gene cooption may have played a role in major episodes of adaptive change in multicellular organisms, and it appears to be an important mechanism for generating morphological and physiological diversity (1–5). Gene duplication may be an especially important facilitator of cooptive evolution (6). This is well illustrated in the vertebrate globin gene superfamily, because there are several well-documented cases where paralogous gene copies have acquired distinct physiological functions and/or patterns of expression (7–11).

Globins are ancient proteins that are present in each of the three domains of life (11–13). Throughout the 20th century, myoglobin (Mb; an O_2 storage protein in muscle) and hemoglobin (Hb; an O_2 transport protein in red blood cells) were the only known globin proteins in vertebrates (8, 14). Early in the 21st century, comparative genomic studies revealed a surprising diversity of novel globin genes in vertebrates, including neuroglobin (Ngb) (15), cytoglobin (Cygb) (16–18), globin-E (GbE) (19), globin-X (GbX) (20), and globin-Y (GbY) (21). The discovery of these novel globin genes has motivated experimental studies to elucidate their physiological functions and evolutionary studies to assess their phylogenetic affinities and taxonomic distributions (22–28).

Phylogenetic studies have revealed that vertebrate globins fall into two distinct clades. One clade contains GbX and Ngb, two highly divergent genes, which appear to be more closely related to annelid intracellular globins than to any other vertebrate globins (20, 21). The other clade contains a set of genes that are products of vertebrate-specific duplication events: Cygb, GbE, GbY, Mb, and the Hbs of jawed vertebrates (gnathostomes) and jawless fish (cyclostomes, represented by lampreys and hagfish) (20–22, 28). The monophyly of these vertebrate-specific globins is well supported (20, 21), but phylogenetic relationships within this group remain highly uncertain.

Because the passive diffusion of O_2 in blood plasma is not generally sufficient to meet the metabolic demands of large, active animals, the evolution of Hb-mediated blood–O_2 transport represented a key physiological innovation in vertebrate life that opened up new opportunities for the evolution of aerobic metabolism. In gnathostomes, Hb is a tetrameric protein assembled from two α-chain and two β-chain subunits. The progenitors of the α- and β-globin gene families arose via tandem duplication of an ancestral, single-copy globin gene approximately 450–500 mya, after the gnathostome common ancestor diverged from jawless fishes (29–31). In the $\alpha_2\beta_2$ Hb tetramers of most extant gnathostomes, the cooperativity of O_2 binding stems from an oxygenation-linked transition in quaternary structure. The origin of cooperativity was preceded by the gene duplication that gave rise to structurally distinct α- and β-chain subunits (11, 30, 32). By contrast, in the Hbs of extant cyclostomes, cooperativity of O_2 binding stems from oxygenation-linked dissociation of multimers into ligated monomers (33–39). For this reason, cyclostome Hbs have been considered "... a transition stage between invertebrate and vertebrate hemoglobins" (40). Phylogenetic studies of vertebrate globins have presented tree topologies that are not consistent with a single origin of O_2 transport Hbs (22, 28, 41). However, incomplete sampling of taxa and gene lineages has not permitted any definitive conclusions.

Here we report a comprehensive phylogenetic reconstruction of the vertebrate globin gene superfamily that includes representatives from each of the major lineages of gnathostomes as well as cyclostomes. Results of this analysis revealed that the erythroid Hbs of cyclostomes and gnathostomes are not orthologous proteins. Instead, the functionally similar O_2 transport proteins were coopted from phylogenetically distinct and anciently diverged globin protein precursors. This represents an example of "cooptive convergence," where paralogous members of the same gene family independently evolve the same spe-

Author contributions: F.G.H., J.C.O., and J.F.S. designed research; performed research; analyzed data; and wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: jstorz2@unl.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1006756107/-/DCSupplemental.

cialization of function in different lineages. After being pressed into service as O_2 transport proteins, the two paralogous globins independently evolved similar biochemical properties in extant cyclostomes and gnathostomes.

The Hbs of cyclostome and gnathostome vertebrates are encapsulated in red blood cells and both proteins provide a highly efficient means of O_2 transport from the respiratory surfaces to the cells of metabolizing tissues while also contributing to the transport of CO_2 back to the gas exchange organs. The efficiency of both Hbs as O_2 transport proteins stems from subunit–subunit interactions (homotropic effects), which account for the cooperativity of O_2 binding, and the deoxygenation-linked binding of allosteric ligands (heterotropic effects), which provides a mechanism for the cellular regulation of Hb–O_2 affinity. In the Hbs of cyclostomes and gnathostomes, cooperativity and allosteric regulation are made possible by oxygenation-linked changes in quaternary structure (42). Thus, the O_2 transport Hbs of both taxa convergently evolved distinct forms of both homotropic and heterotropic cooperative effects from different ancestral protein monomers that lacked cooperativity.

Results

Description of Data. We estimated phylogenetic relationships among all vertebrate-specific members of the globin protein superfamily (Cygb, GbE, GbY, Mb, and Hbs), with special attention to the relationship between cyclostome and gnathostome Hbs. We used a set of vertebrate Ngb sequences to root the tree. To compile the globin sequence dataset, we interrogated the genome assemblies of nine gnathostome vertebrates and used bioinformatic tools to annotate the entire globin gene repertoire of each species. These nine species included representatives of all major gnathostome lineages present in the genome databases (teleost fish, amphibians, squamate reptiles, birds, and mammals). We compiled additional sequences from cartilaginous fish and cyclostomes. When possible, we included more than one species per lineage. Our sample included globin sequences from three cartilaginous fish (red stingray, gummy houndshark, and Port Jackson shark), three teleost fish (medaka, pufferfish, and zebrafish), one amphibian (western clawed frog), one squamate reptile (green anole lizard), two birds (chicken and zebra finch), two mammals (human and platypus), and 12 sequences of functional Hbs from three different cyclostome species: sea lamprey (5 paralogous sequences), Arctic lamprey (3 paralogous sequences), and hagfish (4 paralogous sequences), which cover the two extant cyclostome subclasses Myxini and Hyperoartia. We also retrieved a set of Ngb outgroup sequences from a representative set of gnathostome taxa. A complete description of all sequences used is included as *SI Appendix*, *SI Materials and Methods*, and *SI Appendix*, Table S1.

Most of the gnathostome species included in this study possess multiple paralogous copies of α- and β-like globin genes. Because the monophyly of the α- and β- globin gene families has been well established (30, 31), we only included a representative subset of the α- and β-globins from each species in our analyses.

Phylogenetic Relationships Among Vertebrate Globins. Our primary aims were to reconstruct the globin gene repertoire of the vertebrate common ancestor and to clarify the relationship between cyclostome and gnathostome Hbs. It was traditionally assumed that Hb and Mb originated via duplication of an ancestral, single-copy globin gene before the cyclostome/gnathostome divergence, such that each of these two vertebrate lineages inherited orthologous copies of the same "proto-Hb" gene (11, 30, 32, 40, 43). Under this scenario, we would expect a phylogeny in which the Hbs of cyclostomes are sister to the clade of gnathstome α- and β-Hb genes: [Mb(cyclostome Hb, gnathostome Hb)]. Contrary to this expectation, our maximum likelihood and Bayesian analyses supported a phylogeny in which cyclostome Hb was sister to Cygb, with maximum likelihood bootstrap support of 70% and Bayesian posterior probability of 1.00 (Fig. 1). Cyclostome Hbs, Cygb, GbE, GbYs, and Mb were all placed in strongly supported monophyletic groups, with maximum likelihood bootstrap support values that ranged from 96% to 100% and Bayesian posterior probabilities ≥0.99. Our phylogeny reconstructions also grouped the vertebrate-specific globins into four distinct clades: (*i*) cyclostome Hb + Cygb, (*ii*) Mb + GbE, (*iii*) GbY, and (*iv*) the α- and β-chain Hbs of gnathostomes (this latter clade is sister to the other three clades of vertebrate-specific globins) (Fig. 1).

We performed a comprehensive sensitivity analysis to evaluate how the phylogenetic results were affected by the use of different alignment algorithms, the use of different amino acid substitution models, and the use of different outgroup sequences (e.g., vertebrate GbX or globins from basal chordates such as the sea squirt, *Ciona intestinalis*). To do this, we performed phylogenetic searches for 10 alternative alignments of our sequences under three different models of amino acid substitution. In each of these different analyses, vertebrate globins consistently fell into the four main clades described above, and cyclostome Hb was invariably placed as the sister group to gnathostome Cygb. The bootstrap support value for the node joining cyclostome Hb and Cygb ranged from 48% to 70% among maximum likelihood analyses, whereas posterior probabilities for the same node were far less variable, ranging from 0.99 to 1.00. In all analyses, the trees depicting a sister relationship between cyclostome Hb and gnathostome Cygb had uniformly higher likelihood scores than any of the alternative topologies (*SI Appendix*, Table S2). Finally, we added vertebrate Globin X and *Ciona* globins as additional outgroup sequences, and, again, the tree depicting a sister relationship between cyclostome Hb and gnathostome Cygb had a higher likelihood score than any of the alternatives (*SI Appendix*, Table S3). Full results of the sensitivity analysis are provided in *SI Appendix*, *SI Results*.

The phylogeny reconstruction shown in Fig. 1 provides the basis for two important conclusions: (*i*) precursors of the four main globin gene lineages were all present in the common ancestor of extant vertebrates; and (*ii*) the Hbs of cyclostomes and gnathostomes did not descend from the same ancestral protein in the cyclostome/gnathostome common ancestor. Instead, the cyclostome Hbs are orthologous to the hexacoordinate Cygbs of gnathostome vertebrates. These results are not congruent with any of the previously hypothesized relationships among vertebrate globins. The traditional view regarding the orthology of cyclostome and gnathostome Hbs (11, 30, 32) was based on phylogeny reconstructions that did not include Cygb or other more recently discovered members of the globin protein superfamily (*SI Appendix*, Fig. S1*A*). Alternative phylogenetic relationships have been suggested by more recent work, which included a wider coverage of vertebrate globin diversity. For example, trees presented by Burmester et al. (22, 28) depict a close relationship between GbE and Cygb and a basal position for cyclostome Hbs (*SI Appendix*, Fig. S1*B*). This phylogeny implies that either the O_2 transport functions of cyclostome and gnathostome Hbs were coopted in parallel from a hexacoordinate ancestral state, or alternatively, the O_2 transport function evolved once and was secondarily lost in the lineage that gave rise to the remaining gnathostome-specific globins. Finally, Katoh and Miyata (41) presented a tree in which Cygb was sister to the cyclostome Hbs, and Mb was the most basal of the vertebrate-specific globins (*SI Appendix*, Fig. S1*C*).

These hypotheses make mutually exclusive predictions regarding the evolutionary origins of erythroid O_2 transport Hbs in vertebrates, and these predictions can be tested statistically by using phylogenetic topology tests (44–46). Under the "single cooption" hypothesis (Fig. 2*A*), the Hbs of cyclostomes and gnathostomes descend from the same ancestral precursor protein, and hence, the O_2 transport function evolved only once in the gnathostome/cyclostome common ancestor. This hypothesis predicts that the Hbs

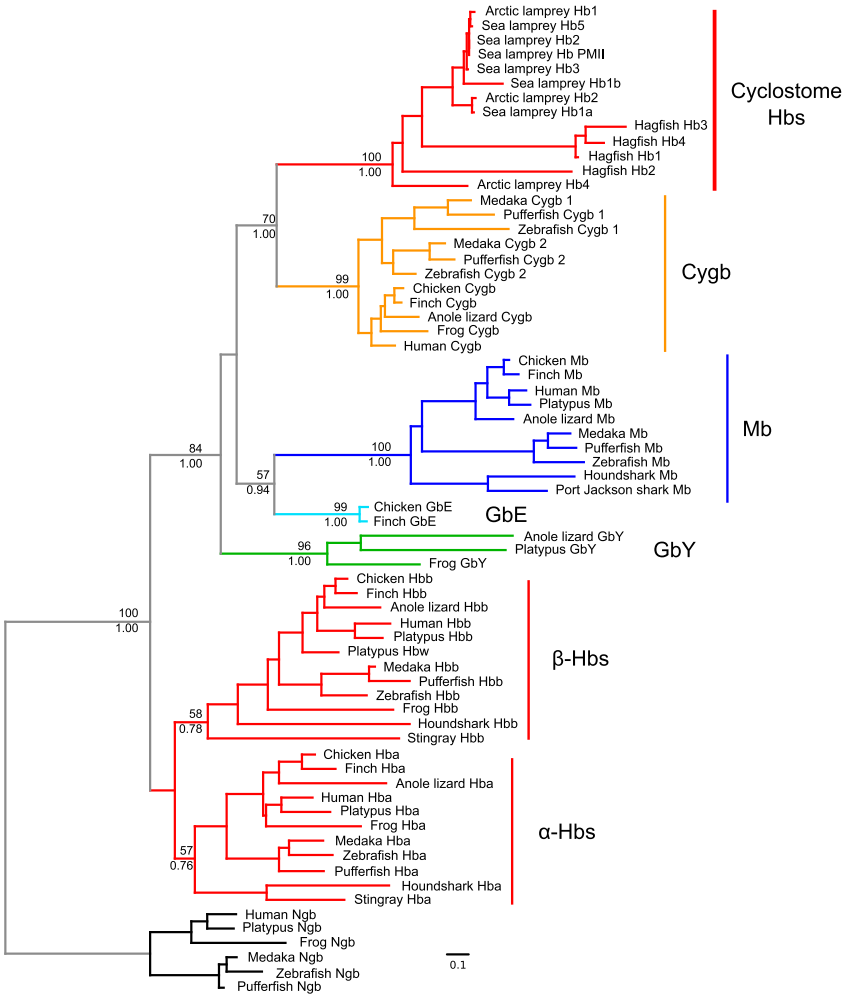


Fig. 1. Maximum likelihood phylogram describing relationships among the Cygb, GbE, GbY, cyclostome Hb, gnathostome α- and β-Hb, and Mb genes of vertebrates. Ngb sequences were included to root the tree. Numbers above the nodes correspond to maximum likelihood bootstrap support values, and those below the nodes correspond to Bayesian posterior probabilities.

of cyclostomes and gnathostomes should form a monophyletic group to the exclusion of all other vertebrate-specific globins, as shown in Fig. 2*A*. If the cyclostome Hbs were the basal vertebrate-specific globin (a "parallel cooption or single cooption/secondary loss" scenario) (22, 28), cyclostome Hbs would be expected to be sister to a clade that contains the gnathostome α- and β-Hbs, Cygb, GbE, GbY, and Mb, as shown in Fig. 2*B*. Finally, if cyclostome and gnathostome Hbs are products of different duplication events (a "convergent cooption" scenario), as suggested by our results and those of Katoh and Miyata (41), we would expect cyclostome Hbs

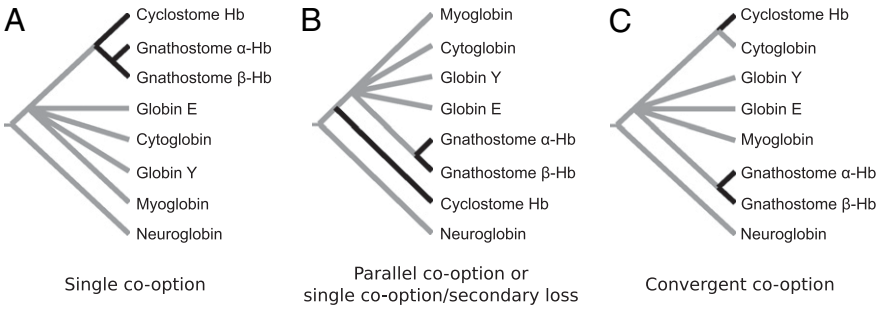


Fig. 2. Alternative hypotheses regarding the phylogenetic relationships between gnathostome and cyclostome Hbs. Under the single cooption hypothesis (*A*), the Hbs of cyclostomes and gnathostomes derive from a proto-Hb precursor protein that acquired an O_2 transport function in the vertebrate common ancestor. Under the parallel cooption or single cooption/secondary loss hypothesis (*B*), the O_2 transport function evolved independently in both Hb lineages, or, alternatively, the function was ancestral and was then secondarily lost in the remaining gnathostome-specific globins. Finally, under the convergent cooption hypothesis (*C*), the O_2 transport function evolved independently in the Hbs of cyclostomes and gnathostomes. Our results favor the convergent cooption hypothesis.

and Cygbs to appear as sister lineages, as shown in Fig. 2*C*. Results of constrained searches favored the independent origin of cyclostome and gnathostome Hbs in all cases (Fig. 2*C* and *SI Appendix*, Table S2). The parametric bootstrapping tests (46, 47) were highly significant in all cases, favoring the convergent cooption scenario ($P \leq 0.001$), and the Shimodaira-Hasegawa and approximately unbiased topology tests lacked power to distinguish among the three hypotheses.

Previous studies have suggested that GbE is more closely related to Cygb than to any other globin (26, 28). Given that GbE has thus far been found only in birds, it was hypothesized to derive from a bird-specific duplication. By contrast, our phylogeny indicates that GbE is more closely related to Mb than to any other globin, and the fact that GbE and Mb are located on the same chromosome in birds is consistent with the phylogenetic results. It thus appears that Mb and GbE derive from a duplication event that predated the gnathostome radiation, and subsequently, the GbE ortholog was independently lost in all gnathostome lineages other than birds. We postulate a similar scenario for GbY, as this gene was probably present in the ancestor of all extant vertebrates and was independently lost in all lineages other than amphibians (as represented by *Xenopus*), squamate reptiles (as represented by *Anolis*), and monotreme mammals (as represented by the platypus). Interestingly, the multiple Hbs of lampreys and hagfish do not form reciprocally monophyletic groups (Fig. 1). The phylogenetic patterns indicate that both lineages inherited at least two Hb paralogs from the cyclostome common ancestor (approximately 450 mya) (48, 49), and the globin gene repertoires of lampreys and hagfish were then further expanded by subsequent rounds of lineage-specific gene duplication and divergence.

Discussion

Vertebrate-specific globins can be grouped into four distinct lineages, as represented by (*i*) cyclostome Hb + Cygb, (*ii*) Mb + GbE, (*iii*) GbY, and (*iv*) the α- and β-chain Hbs of gnathostomes. The common ancestor of extant vertebrates possessed a globin gene repertoire that included progenitors of each of these four distinct gene lineages. Representatives of the first three of these lineages have not been found in cyclostomes, whereas gnathostomes appear to have retained representatives of all four paralogous gene lineages. Subsequent gene duplications and gene losses have occurred in different gnathostome lineages, as illustrated by the independent loss of GbE in all gnathostomes other than birds. These results demonstrate that variation in the globin gene repertoire among extant vertebrates can be attributed to differential retention and loss of an ancestral gene set that was inherited from the vertebrate common ancestor roughly 600 mya in the Cambrian Period.

Cooptive Convergence of Protein Function. Beyond a certain body size threshold, simple diffusion of O_2 in blood plasma is generally not sufficient to meet the metabolic demands of animal life. Here we report the surprising discovery that similar physiological problems have called forth similar solutions in different lineages during the basal radiation of vertebrates. We discovered that the ancestors of cyclostome and gnathostome vertebrates independently invented erythroid-specific O_2 transport Hbs as a means of enhancing blood–O_2 transport (Fig. 3). In this context, applying the name "Hb" to both proteins simply denotes functional analogy and not homology (8, 11, 28). Although cyclostomes and gnathostomes make use of functionally similar respiratory pigments for blood–gas transport, the superficial similarities in protein function do not reflect continuity of inheritance from a common ancestral protein. Our phylogeny reconstruction indicates that cyclostome Hbs are most closely related to gnathostome Cygb, a hexacoordinate globin protein that is predominantly expressed in the cytoplasm of cells that are actively engaged in the production of extracellular matrix components in visceral organs. The protein may also play a role in intracellular signaling pathways or other functions related to cellular O_2 metabolism (22–28). Some of the functionally similar features related to homo- and heterotropic interactions have a different structural basis in cyclostome and gnathostome Hbs (38, 39), as might be expected if the functions were coopted and modified from different precursor proteins (i.e., different ancestral states). In both cases, multisubunit structures provided the basis for cooperative O_2 binding by coupling the effects of ligand binding at individual subunits with interactions between subunits, but differences in numerous structural details belie their independent origins.

In contrast to the tetrameric Hbs of most gnathostomes, the Hbs of cyclostomes exist as monomers in the oxygenated state and self-associate into dimers or tetramers upon deoxygenation (33–39). This oxygenation-linked reversible aggregation accounts for a modest degree of cooperativity, and the release of Bohr protons upon dissociation into monomers provides a mechanism of allosteric regulation. Heterotetrameric Hbs of the hagfish *Eptatretus burger* exhibit significant cooperativity (50), and evidence for similar subunit interactions have been documented in

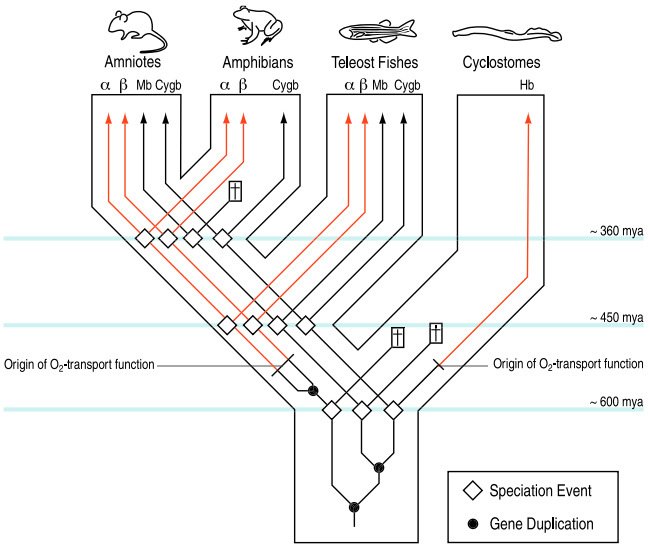


Fig. 3. An evolutionary model describing the independent evolution of erythroid-specific O_2 transport Hbs in gnathostomes and cyclostomes. The gnathostomes are here represented by amniotes, amphibians, and teleost fish. According to this model, the duplication of an ancestral, single-copy globin gene in the stem lineage of vertebrates produced one descendant gene lineage that eventually gave rise to the α- and β-Hbs of gnathostomes (left branch) and another gene lineage that gave rise to the common ancestor of Mb, Cygb, and cyclostome Hb (right branch). In the latter gene lineage, a subsequent duplication before the vertebrate radiation gave rise to Mb (which was secondarily lost in cyclostomes) and Cygb (= cyclostome Hb). In the gnathostome and cyclostome Hb gene lineages, independent origins of O_2 transport functions are denoted by orange lines. In the case of gnathostome Hb, cooperativity of O_2 binding stems from oxygenation-linked transitions in quaternary structure of an $\alpha_2\beta_2$ heterotetramer. The origin of cooperativity and allosteric regulation was preceded by duplication and divergence of the proto α- and β-globin genes. In the case of cyclostome Hb, cooperativity stems from oxygenation-linked dissociation of multimers into ligated monomers. The O_2 transport function of gnathostome Hb preceded the divergence of cartilaginous fish from the ancestor of teleosts and tetrapods approximately 500 mya (48, 78), and the O_2 transport function of cyclostome Hb preceded the divergence between representatives of the two extant cyclostome subclasses, Myxini and Hyperoartia, approximately 450 mya (49). Thus, the convergent evolution of erythroid-specific O_2 transport Hbs appears to have occurred in the early Paleozoic (approximately 450–600 mya), after the split between gnathostomes and cyclostomes, but before the split between cartilaginous fish and the ancestor of teleosts and tetrapods and before the split between hagfish and lamprey. Estimated divergence dates are taken from Hedges (78).

multimeric Hb isoforms of the hagfish *Myxine glutinosa* (37). The formation of heteromultimers composed of unlike subunits appears superficially similar to the $\alpha_2\beta_2$ heterotetramers of most gnathostomes, but the oxygenation-linked transition in quaternary structure is completely different. The intersubunit contacts of heterotetrameric gnathostome Hbs primarily involve the C, G, and H helices of the globin chain subunits (51), whereas the contact surfaces of the deoxygenated, homodimeric cyclostome Hbs involve the E helix and the AB corner, such that the heme groups are in almost direct contact (52–55). The heme–heme interactions of cyclostome Hbs are intriguingly similar to those of the homodimer of Cygb (23, 38, 39, 56, 57), an observation that makes sense in light of our inferred phylogenetic relationship between these two globin proteins (Fig. 1).

The convergent or parallel evolution of a given trait in different phylogenetic lineages can often be interpreted as evidence that the trait confers an adaptive advantage. For a globin protein with an O_2 transport function (as opposed to O_2 storage, O_2 scavenging, or O_2 sensing functions), cooperativity is advantageous because it permits rapid and efficient O_2 unloading over a relatively narrow range of blood–O_2 tensions. Moreover, cooperativity permits O_2 unloading at higher partial pressures of O_2 than is possible in the absence of cooperativity, thereby maintaining a pressure gradient between capillary plasma and the tissue mitochondria. The pH dependence of Hb–O_2 affinity (Bohr effect) is advantageous in active animals because it increases the efficiency of O_2 delivery to metabolizing tissues (58, 59). This may explain why the magnitude of the Bohr effect is substantially greater in the Hbs of lampreys than in the generally less active hagfish (39, 60–63).

Conclusion

The ancestors of extant cyclostomes and gnathostomes independently evolved O_2 transport globin proteins by exploiting different mechanisms of oxygenation-linked conformational change in a multisubunit structure. In both cases, the underlying genes also independently evolved erythroid-specific expression. In the Hbs of both gnathostomes and cyclostomes, cooperative O_2 binding is made possible by coupling the effects of ligand binding at individual subunits and the interactions between subunits in the quaternary structure. This example of convergent evolution of protein function provides an impressive demonstration of the ability of natural selection to cobble together complex design solutions by tinkering with different variations of the same basic protein scaffold.

Materials and Methods

Phylogenetic Inference. Because the goal of this study was to estimate phylogenetic relationships among the Cygb, GbE, GbY, Hb, and Mb genes of vertebrates, we included Ngb sequences from human, platypus, frog, medaka, tetraodon, and zebrafish to root the tree. Previous results indicate that Ngb is an appropriate outgroup because it is more closely related to annelid intracellular globins than to any other vertebrate globin (20, 28). Phylogenetic relationships were estimated using maximum likelihood and Bayesian methods. Because the use of different sequence alignments and substitution models may influence the results of phylogenetic analyses (64, 65), we conducted a comprehensive sensitivity analysis. Specifically, we aligned sequences using 10 alternative methods, and for each resulting alignment, we performed maximum likelihood and Bayesian analyses using two different substitution models. Briefly, we aligned sequences using Dialign (66), Kalign2 (67), the E-INS-i, G-INS-i, and L-INS-i strategies from Mafft v6.17 (68), Muscle v3.5 (69), Prank (70), Probalign (71), Probcons (72), and PROMALS3d (73). A data file containing the complete set of sequence alignments is provided in the *SI Appendix* and Dataset S1. Maximum likelihood searches were performed under JTT (74), LG (75), and mixed models, and Bayesian searches were performed under the JTT (74) and mixed models.

We report primary results that were based on the Muscle alignment and the mixed model of amino acid substitution. We report all other results in *SI Appendix*, Table S1. Maximum likelihood searches were implemented in Treefinder version October 2008 (76), and support for the nodes was evaluated with 1,000 bootstrap pseudoreplicates. Bayesian analyses were conducted using MrBayes version 3.1.2 (77), setting two independent runs of four simultaneous chains for 10,000,000 generations, sampling every 2,500 generations, and using default priors. Once convergence was verified, support for the nodes and parameter estimates were derived from a majority rule consensus of the last 2,500 trees.

Hypothesis Testing. We compared alternative hypotheses using the Shimodaira-Hasegawa (45), approximately unbiased (44), and parametric bootstrapping tests (46, 47). In the case of parametric bootstrapping, for each simulated data set, we calculated the difference in likelihood score, Δ, between the null hypothesis maximum likelihood topology and the alternative hypothesis maximum likelihood topology. Using an α level of 0.01, the null hypothesis maximum likelihood topology was rejected if $\geq$99% of the simulation-based Δ values exceeded the observed value. All tests were carried out in Treefinder.

ACKNOWLEDGMENTS. We thank Z. Cheviron, M. Goodman, A. Runck, and two anonymous reviewers for helpful comments and suggestions. This work was funded by grants to J.F.S. from the National Science Foundation and National Institutes of Health/ National Heart, Lung, and Blood Institute and grants to J.C.O. from the Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT 11080181), the Programa Bicentenario en Ciencia y Tecnología (PSD89), and the Concurso Estadía Jóvenes Investigadores en el Extranjero from the Universidad Austral de Chile. This work made use of resources provided by the Computational Biology Service Unit from Cornell University, which is partially funded by Microsoft Corporation.

1. Eizinger A, Jungblut B, Sommer RJ (1999) Evolutionary change in the functional specificity of genes. *Trends Genet* 15:197–202.
2. Ganfornina MD, Sánchez D (1999) Generation of evolutionary novelty by functional shift. *Bioessays* 21:432–439.
3. Carroll SB (2001) Chance and necessity: The evolution of morphological complexity and diversity. *Nature* 409:1102–1109.
4. True JR, Carroll SB (2002) Gene co-option in physiological and morphological evolution. *Annu Rev Cell Dev Biol* 18:53–80.
5. Piatigorsky J (2007) *Gene Sharing and Evolution: The Diversity of Protein Functions* (Harvard University Press, Cambridge, MA).
6. Conant GC, Wolfe KH (2008) Turning a hobby into a job: How duplicated genes find new functions. *Nat Rev Genet* 9:938–950.
7. Hardison R (2001) *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Managment*, eds Steinberg MH, Forget BG, Higgs DR, Nagel RL (Cambridge University Press, Cambridge, UK), pp 95–115.
8. Hardison R (1998) Hemoglobins from bacteria to man: Evolution of different patterns of gene expression. *J Exp Biol* 201:1099–1117.
9. Opazo JC, Hoffmann FG, Storz JF (2008) Genomic evidence for independent origins of β-like globin genes in monotremes and therian mammals. *Proc Natl Acad Sci USA* 105: 1590–1595.
10. Hoffmann FG, Storz JF, Gorr TA, Opazo JC (2010) Lineage-specific patterns of functional diversification in the α- and β-globin gene families of tetrapod vertebrates. *Mol Biol Evol* 27:1126–1138.
11. Goodman M, et al. (1988) An evolutionary tree for invertebrate globin sequences. *J Mol Evol* 27:236–249.
12. Vinogradov SN, Moens L (2008) Diversity of globin function: Enzymatic, transport, storage, and sensing. *J Biol Chem* 283:8773–8777.
13. Vinogradov SN, et al. (2007) A model of globin evolution. *Gene* 398:132–142.
14. Hardison R (2005) *Encyclopedia of Life Sciences* (John Wiley & Sons, Ltd, Chichester, UK).
15. Burmester T, Weich B, Reinhardt S, Hankeln T (2000) A vertebrate globin expressed in the brain. *Nature* 407:520–523.
16. Kawada N, et al. (2001) Characterization of a stellate cell activation-associated protein (STAP) with peroxidase activity found in rat hepatic stellate cells. *J Biol Chem* 276:25318–25323.
17. Trent JTIII, 3rd, Hargrove MS (2002) A ubiquitously expressed human hexacoordinate hemoglobin. *J Biol Chem* 277:19538–19545.
18. Burmester T, Ebner B, Weich B, Hankeln T (2002) Cytoglobin: A novel globin type ubiquitously expressed in vertebrate tissues. *Mol Biol Evol* 19:416–421.
19. Kugelstadt D, Haberkamp M, Hankeln T, Burmester T (2004) Neuroglobin, cytoglobin, and a novel, eye-specific globin from chicken. *Biochem Biophys Res Commun* 325:719–725.
20. Roesner A, Fuchs C, Hankeln T, Burmester T (2005) A globin gene of ancient evolutionary origin in lower vertebrates: Evidence for two distinct globin families in animals. *Mol Biol Evol* 22:12–20.
21. Fuchs C, Burmester T, Hankeln T (2006) The amphibian globin gene repertoire as revealed by the *Xenopus* genome. *Cytogenet Genome Res* 112:296–306.
22. Burmester T, et al. (2004) Neuroglobin and cytoglobin: Genes, proteins and evolution. *IUBMB Life* 56:703–707.
23. Fago A, Hundahl C, Malte H, Weber RE (2004) Functional properties of neuroglobin and cytoglobin. Insights into the ancestral physiological roles of globins. *IUBMB Life* 56:689–696.
24. Pesce A, et al. (2004) Reversible hexa- to penta-coordination of the heme Fe atom modulates ligand binding properties of neuroglobin and cytoglobin. *IUBMB Life* 56:657–664.

25. Pesce A, et al. (2002) Neuroglobin and cytoglobin. Fresh blood for the vertebrate globin family. *EMBO Rep* 3:1146–1151.
26. Hankeln T, et al. (2005) Neuroglobin and cytoglobin in search of their role in the vertebrate globin family. *J Inorg Biochem* 99:110–119.
27. Hankeln T, Burmester T (2008) *The Smallest Biomolecules: Diatomics and Their Interactions with Heme Proteins*, ed Gosh A (Elsevier B.V., Amsterdam, The Netherlands), pp 203–218.
28. Burmester T, Hankeln T (2009) What is the function of neuroglobin? *J Exp Biol* 212: 1423–1428.
29. Czelusniak J, et al. (1982) Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. *Nature* 298:297–300.
30. Goodman M, Moore GW, Matsuda G (1975) Darwinian evolution in the genealogy of haemoglobin. *Nature* 253:603–608.
31. Goodman M, Miyamoto MM, Czelusniak J (1987) *Molecules and Morphology in Evolution: Conflict or Compromise?* ed Patterson C (Cambridge University Press, Cambridge, UK), pp 140–176.
32. Coates ML (1975) Hemoglobin function in the vertebrates: An evolutionary model. *J Mol Evol* 6:285–307.
33. Rumen NM, Love WE (1963) The six hemoglobins of the sea lamprey (*Petromyzon marinus*). *Arch Biochem Biophys* 103:24–35.
34. Li SL, Riggs A (1970) The amino acid sequence of hemoglobin V from the lamprey, *Petromyzon marinus*. *J Biol Chem* 245:6149–6169.
35. Andersen ME (1971) Sedimentation equilibriujm experiments on the self-assocation of hemoglobin from the lamprey *Petromyzon marinus*. A model for oxygen transport in the lamprey. *J Biol Chem* 246:4800–4806.
36. Brittain T, Wells RM (1986) Characterization of the changes in the state of aggregation induced by ligand binding in the hemoglobin system of a primitive vertebrate, the hagfish *Eptatretus cirrhatus*. *Comp Biochem Physiol Comp Physiol* 85: 785–790.
37. Fago A, Weber RE (1995) The hemoglobin system of the hagfish *Myxine glutinosa*: Aggregation state and functional properties. *Biochim Biophys Acta* 1249:109–115.
38. Qiu Y, Maillett DH, Knapp J, Olson JS, Riggs AF (2000) Lamprey hemoglobin. Structural basis of the bohr effect. *J Biol Chem* 275:13517–13528.
39. Fago A, et al. (2001) Hagfish hemoglobins: Structure, function, and oxygen-linked association. *J Biol Chem* 276:27415–27423.
40. Müller G, Fago A, Weber RE (2003) Water regulates oxygen binding in hagfish (*Myxine glutinosa*) hemoglobin. *J Exp Biol* 206:1389–1395.
41. Katoh K, Miyata T (2002) Cyclostome hemoglobins are possibly paralogous to gnathostome hemoglobins. *J Mol Evol* 55:246–249.
42. Weber RE, Fago A (2004) Functional adaptation and its molecular basis in vertebrate hemoglobins, neuroglobins and cytoglobins. *Respir Physiol Neurobiol* 144:141–159.
43. Tiplady B, Goodman M (1977) Primitive haemoglobin. *J Mol Evol* 9:343–347.
44. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51:492–508.
45. Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16:1114–1116.
46. Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) *Molecular Systematics*, eds Hillis DM, Moritz C, Mable B (Sinauer, Sunderland, MA) 2nd Ed pp 407–514.
47. Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. *Syst Biol* 49:652–670.
48. Blair JE, Hedges SB (2005) Molecular phylogeny and divergence times of deuterostome animals. *Mol Biol Evol* 22:2275–2284.
49. Kuraku S, Kuratani S (2006) Time scale for cyclostome evolution inferred with a phylogenetic diagnosis of hagfish and lamprey cDNA sequences. *Zoolog Sci* 23: 1053–1064.
50. Bannai S, Sugita Y, Yoneyama Y (1972) Studies on hemoglobin from the hagfish Epatatretus burgeri. *J Biol Chem* 247:505–510.
51. Perutz MF, Fermi G, Luisi B, Shaanan B, Liddington RC (1987) Stereochemistry of cooperative mechanisms in hemoglobin. *Cold Spring Harb Symp Quant Biol* 52: 555–565.
52. Honzatko RB, Hendrickson WA (1986) Molecular models for the putative dimer of sea lamprey hemoglobin. *Proc Natl Acad Sci USA* 83:8487–8491.
53. Riggs AF (1998) Self-association, cooperativity and supercooperativity of oxygen binding by hemoglobins. *J Exp Biol* 201:1073–1084.
54. Heaslet HA, Royer WEJ, Jr (1999) The 2.7 A crystal structure of deoxygenated hemoglobin from the sea lamprey (*Petromyzon marinus*): Structural basis for a lowered oxygen affinity and Bohr effect. *Structure* 7:517–526.
55. Mito M, et al. (2002) Crystal structures of deoxy- and carbonmonoxyhemoglobin F1 from the hagfish *Eptatretus burgeri*. *J Biol Chem* 277:21898–21905.
56. Sugimoto H, et al. (2004) Structural basis of human cytoglobin for ligand binding. *J Mol Biol* 339:873–885.
57. Makino M, et al. (2006) High-resolution structure of human cytoglobin: Identification of extra N- and C-termini and a new dimerization mode. *Acta Crystallogr D Biol Crystallogr* 62:671–677.
58. Giardina B, Mosca D, De Rosa MC (2004) The Bohr effect of haemoglobin in vertebrates: An example of molecular adaptation to different physiological requirements. *Acta Physiol Scand* 182:229–244.
59. Jensen FB (2004) Red blood cell pH, the Bohr effect, and other oxygenation-linked phenomena in blood O2 and CO2 transport. *Acta Physiol Scand* 182:215–227.
60. Nikinmaa M (1993) Haemoglobin function in intact *Lampetra fluviatilis* erythrocytes. *Respir Physiol* 91:283–293.
61. Fago A, Weber RE (1998) *The Biology of Hagfishes*, eds Jørgensen JM, Lomholt JP, Weber RE, Malte H (Chapman & Hall, London), pp 321–329.
62. Fago A, Malte H, Dohn N (1999) Bicarbonate binding to hemoglobin links oxygen and carbon dioxide transport in hagfish. *Respir Physiol* 115:309–315.
63. Jensen FB (1994) Haemoglobin H+ equilibria in lamprey (Lampetra fluviatilis) and hagfish (Myxine glutinosa). *J Exp Biol* 202:1963–1968.
64. Sullivan J, Joyce P (2005) Model selection in phylogenetics. *Annu Rev Ecol Evol Syst* 36: 445–466.
65. Wong KM, Suchard MA, Huelsenbeck JP (2008) Alignment uncertainty and genomic analysis. *Science* 319:473–476.
66. Morgenstern B (2004) DIALIGN: Multiple DNA and protein sequence alignment at BiBiServ. *Nucleic Acids Res* 32 (Web Server issue):W33–W36.
67. Lassmann T, Frings O, Sonnhammer ELL (2009) Kalign2: High-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Res* 37:858–865.
68. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9:286–298.
69. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
70. Löytynoja A, Goldman N (2005) An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci USA* 102:10557–10562.
71. Chikkagoudar S, Roshan U, Livesay D (2007) eProbalign: Generation and manipulation of multiple sequence alignments using partition function posterior probabilities. *Nucleic Acids Res* 35 (Web Server issue):W675–W677.
72. Do CB, Mahabhashyam MSP, Brudno M, Batzoglou S (2005) ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res* 15:330–340.
73. Pei J, Tang M, Grishin NV (2008) PROMALS3D web server for accurate multiple protein sequence and structure alignments. *Nucleic Acids Res* 36 (Web Server issue):W30–W34.
74. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8:275–282.
75. Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* 25:1307–1320.
76. Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: A powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4:18.
77. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
78. Hedges S (2009) *The Timetree of Life*, eds Kumar S, Hedges S (Oxford University Press, Oxford, UK), pp 309–314.

Supporting Information

Material and Methods

Sequence Data. We retrieved the full complement of globin genes from the genome sequences of nine vertebrate taxa, including three teleost fish (medaka, *Oryzias latipes*; pufferfish, *Tetraodon nigroviridis*; and zebrafish, *Danio rerio*), one amphibian (western clawed frog, *Xenopus tropicalis*), one squamate reptile (green anole lizard, *Anolis carolinensis*), two birds (chicken, *Gallus gallus*; and zebra finch, *Taeniopygia guttata*), and two mammals (human, *Homo sapiens*; and platypus, *Ornithorhynchus anatinus*). The complete globin gene repertoire of each species was obtained by means of bioinformatic searches in the Genbank or Ensembl (release 55) databases. We broadened our phylogenetic coverage by adding globin sequences derived from mRNA or protein records from representative cartilaginous fish (class Chondrichthyes), the most basal lineage of extant gnathostomes, and from cyclostomes, the sister group to gnathostome vertebrates. In the case of cartilaginous fish, we obtained α- and β-Hb sequences from the red stingray (*Dasyatis akajei*) and gummy houndshark (*Mustelus antarcticus*), as well as Mb sequences from the latter species and the Port Jackson shark (*Heterodontus portusjacksoni*). In the case of cyclostomes, we included 12 sequences of functional Hbs from three representatives of subclasses Myxini and Hyperoartia: Five paralogous sequences from the sea lamprey (*Petromyzon marinus*, Hyperoartia), three from the Arctic lamprey (*Lethenteron japonicum*, Hyperoartia), and four from the hagfish (*Myxine glutinosa,* Myxini). We did not identify any previously undescribed globin genes in the Ensembl pre-release assembly of the sea lamprey genome. In addition, we included the previously reported globins from the sea squirt, *Ciona intestinalis* (1).

SI Table 1.
Sequences used in this study with the corresponding accession numbers.

Sequence Name	Accesion number	Source
Anole lizard Cygb	ENSACAG00000008394*	Ensembl
Anole lizard GbY	AAWZ01045931*	Ensembl
Anole lizard Hba	ENSACAG00000016421	Ensembl
Anole lizard Hbb	ENSACAG00000012173	Ensembl
Anole lizard Mb	ENSACAG00000016595	Ensembl
Chicken Cygb	NM_001008789	GenBank
Chicken GbE	NM_001008786	GenBank
Chicken Hba	NM_001004376	GenBank
Chicken Hbb	NM_001081704	GenBank
Chicken Mb	XM_416292	GenBank
Frog Cygb	NM_001006869	GenBank
Frog GbX	NP_001011196	GenBank
Frog GbY	BC158411	GenBank
Frog Hba	NM_203529	GenBank
Frog Hbb	NM_203528	GenBank
Frog Ngb	ENSXETG00000027106	Ensembl
Hagfish Hb1	AF156936	GenBank
Hagfish Hb2	AF157494	GenBank
Hagfish Hb3	AF184047	GenBank
Hagfish Hb4	AF184239	GenBank
Houndshark Hba	BAA75399	GenBank
Houndshark Hbb	BAA75400	GenBank
Houndshark Mb	P14399	GenBank
Human Cygb	NM_134268	GenBank
Human Hba	NM_000558	GenBank
Human Hbb	NM_000518	GenBank
Human Mb	NG_007075	GenBank
Human Ngb	NP_067080	GenBank
Lethenteron Hb 1	AB294235	GenBank
Lethenteron Hb 2	AB294236	GenBank
Lethenteron Hb 4	AB294237	GenBank
Medaka Cygb-1	NM_001104767	GenBank
Medaka Cygb-2	NM_001104768	GenBank
Medaka GbX	ENSORLG00000017054	Ensembl
Medaka Hba	ENSORLG00000005267	Ensembl
Medaka Hbb	ENSORLG00000003020	Ensembl
Medaka Mb	ENSORLG00000004130	Ensembl
Medaka Ngb	ENSORLT00000020359	Ensembl

Sequence Name	Accesion number	Source
Platypus GbY	AC203513	GenBank
Platypus Hba	AC203513	GenBank
Platypus Hbb	AC190020	GenBank
Platypus Hbw	AC203513	GenBank
Platypus Mb	XM_001513063	GenBank
Platypus Ngb	XP_001508417	GenBank
Port Jackson shark Mb	P02206	GenBank
Pufferfish Cygb-1	AJ635230	GenBank
Pufferfish Cygb-2	AJ635231	GenBank
Pufferfish GbX	CAG25725	GenBank
Pufferfish Hba	ENSTNIG00000018576	Ensembl
Pufferfish Hbb	ENSTNIG00000012913	Ensembl
Pufferfish Mb	ENSTNIG00000005518	Ensembl
Pufferfish Ngb	CAC59974	GenBank
Sea lamprey Hb PMII	AF248645	GenBank
Sea lamprey Hb1a	P09967	GenBank
Sea lamprey Hb1b	P21197	GenBank
Sea lamprey Hb2	Q9I9I3	GenBank
Sea lamprey Hb3	P09968	GenBank
Sea lamprey Hb5	P02208	GenBank
Sea squirt Glb1	CAD68145	GenBank
Sea squirt Glb2	CAD68146	GenBank
Sea squirt Glb3	CAD68147	GenBank
Sea squirt Glb4	CAD89600	GenBank
Stingray Hba	BAA75249	GenBank
Stingray Hbb	BAA75250	GenBank
Zebra finch Cygb	XM_002195407	GenBank
Zebra finch GbE	XM_002196350	GenBank
Zebra finch Hba	XM_002196096	GenBank
Zebra finch Hbb	XM_002190485	GenBank
Zebra finch Mb	XM_002199380	GenBank
Zebrafish Cygb1	NM_152952	GenBank
Zebrafish Cygb2	NM_001024224	GenBank
Zebrafish GbX	CAG25723	GenBank
Zebrafish Hba	NM_131257	GenBank
Zebrafish Hbb	NM_131020	GenBank
Zebrafish Mb	NM_200586	GenBank
Zebrafish Ngb	NP_571928	GenBank

* These sequences were re-annotated manually.

Results

Sensitivity analysis. Because changes in sequence alignment and substitution model are known to influence the results of phylogenetic analyses (2, 3),we explored sensitivity of our results to variation in alignment method, substitution model, and choice of outgroup. We aligned sequences with 10 alternative methods: Dialign (4), Kalign2 (5), the E-INS-i, G-INS-i, and L-INS-i strategies from Mafft v6.17 (6), Muscle v3.5 (7), Prank (8), Probalign (9), Probcons (10), and PROMALS3d (11). For each alignment we performed maximum likelihood searches under the JTT (12), LG (13), and mixed models of amino acid substitution and Bayesian analyses under the JTT (12), and mixed models of amino acid substitution. Maximum likelihood searches were implemented in Treefinder version October 2008 (14), and support for the nodes was evaluated with 1,000 bootstrap pseudoreplicates. Bayesian analyses were conducted using MrBayes version 3.1.2 (15), setting two independent runs of four simultaneous chains for 10,000,000 generations, sampling every 2,500 generations, and using default priors. Once convergence was verified, support for the nodes and parameter estimates were derived from a majority rule consensus of the last 2,500 trees. In maximum likelihood we used constrained searches to compare the likelihood scores of the ‘single co-option’ hypothesis (Fig. 2A, SI Fig 1A), the ‘parallel co-option or single co-option/secondary loss’ hypothesis (Fig. 2B, SI Fig 1B), and the ‘convergent co-option’ hypothesis (Fig. 2C, SI Fig 1C). Finally, we added vertebrate Globin X sequences and *Ciona* globin sequences to the alignment as additional outgroup sequences. A data file containing the complete set of sequence alignments is provided in the Supporting Information online.

SI Table 2. Results of the sensitivity analysis. Maximum likelihood scores of the best unconstrained tree, and the three competing hypotheses of globin gene family evolution, plus support for the node joining cyclostome hemoglobins with gnathostome cytoglobin. This first set of analyses included all the vertebrate-specific globins, plus six vertebrate neuroglobin sequences as outgroup sequences.

Data: vertebrate-specific globins + Neuroglobin

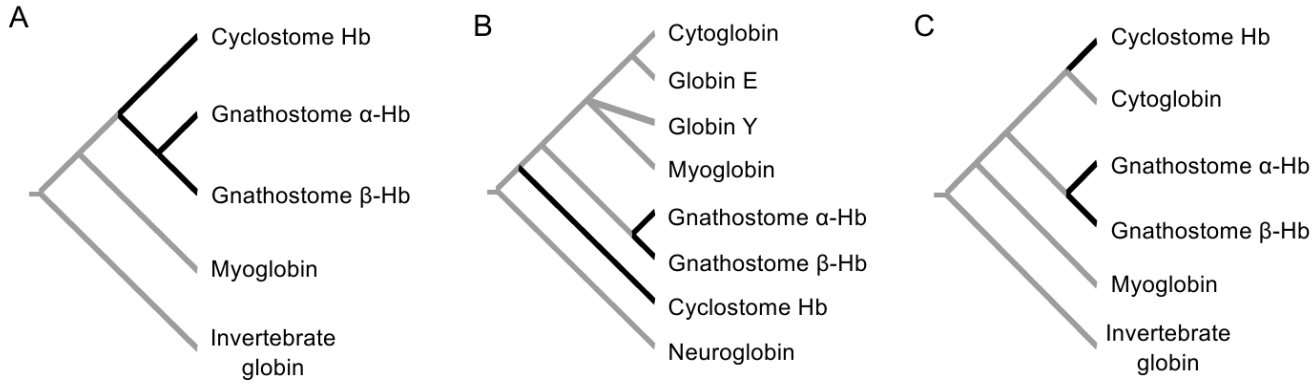
model = JTT		Likelihood scores			Support for the node joining cyclostome Hbs and cytoglobin	
Alignment	best tree	'single co-option'	'parallel co-option or single co-option/secondary loss'	'convergent co-option'	ML bs	MrBayes pp
dialign	-13480.4	-13491.6	-13491.3	-13481.1	63%	1.00
kalign	-13354.1	-13366.1	-13361.7	-13354.1	58%	1.00
mafft_einsi	-13351.2	-13361.6	-13361.0	-13351.2	65%	1.00
mafft_ginsi	-13325.5	-13332.5	-13331.0	-13322.7	66%	1.00
mafft_linsi	-13351.2	-13361.6	-13361.0	-13351.2	64%	1.00
muscle	-13417.7	-13430.3	-13426.8	-13416.7	76%	1.00
prank	-13255.0	-13265.8	-13262.3	-13255.0	73%	1.00
probalign	-13476.1	-13487.4	-13486.1	-13475.6	65%	1.00
probcons	-13433.3	-13445.6	-13442.7	-13433.3	59%	1.00
promal_wPDB	-13512.2	-13522.4	-13520.9	-13512.8	65%	1.00
model = LG						
dialign	-13434.8	-13443.9	-13444.6	-13429.3	54%	--
kalign	-13293.4	-13300.7	-13300.0	-13296.8	< 50%	--
mafft_einsi	-13279.8	-13291.0	-13289.1	-13279.8	57%	--
mafft_ginsi	-13253.7	-13261.1	-13260.3	-13252.3	58%	--
mafft_linsi	-13279.8	-13291.0	-13289.1	-13279.8	54%	--
muscle	-13355.7	-13367.2	-13364.4	-13356.2	67%	--
prank	-13197.1	-13205.3	-13201.3	-13197.1	68%	--
probalign	-13415.2	-13425.5	-13423.6	-13417.1	59%	--
probcons	-13362.0	-13370.5	-13369.1	-13362.0	< 50%	--
promal_wPDB	-13457.0	-13465.8	-13464.4	-13457.0	58%	--
model = mixed						
dialign	-13406.7	-13417.1	-13415.2	-13406.9	57%	1.00
kalign	-13279.1	-13286.0	-13285.2	-13279.1	54%	0.99
mafft_einsi	-13272.0	-13281.6	-13280.7	-13272.4	59%	1.00
mafft_ginsi	-13225.4	-13235.0	-13234.0	-13225.2	59%	1.00
mafft_linsi	-13272.0	-13281.6	-13280.7	-13272.4	55%	0.99
muscle	-13339.6	-13350.9	-13347.8	-13340.1	70%	1.00
prank	-13151.6	-13162.7	-13159.1	-13151.6	70%	0.99
probalign	-13388.9	-13399.6	-13397.1	-13388.9	63%	1.00
probcons	-13351.5	-13357.4	-13357.5	-13349.7	50%	0.99
promal_wPDB	-13428.4	-13437.9	-13436.0	-13428.5	60%	1.00

SI Table 3. Results of the sensitivity analysis. Maximum likelihood scores of the best unconstrained tree, and the three competing hypotheses of globin gene family evolution. This second set of analyses included all the vertebrate-specific globins, plus four vertebrate Globin X sequences, four *Ciona* globin sequences, and the six vertebrate neuroglobin sequences as outgroup sequences.

Data: vertebrate-specific globins + Neuroglobin + Globin X + *Ciona* globins

model = JTT			Likelihood scores	
Alignment	best tree	'single co-option'	'parallel co-option or single co-option/secondary loss'	'convergent co-option'
dialign	-16406.3	-16412.3	-16405.0	-16402.3
kalign	-16279.1	-16284.0	-16281.2	-16278.6
mafft_einsi	-16253.2	-16260.3	-16254.8	-16254.0
mafft_ginsi	-16221.8	-16227.5	-16225.9	-16221.0
mafft_linsi	-16264.2	-16268.7	-16267.4	-16262.4
muscle	-16521.8	-16529.8	-16523.5	-16521.6
prank	-16143.5	-16150.7	-16143.0	-16142.1
probalign	-16513.6	-16521.5	-16516.1	-16514.5
probcons	-16449.4	-16453.5	-16450.6	-16449.4
promal_wPDB	-16440.8	-16448.5	-16442.3	-16440.8
model = LG				
dialign	-16349.3	-16349.9	-16350.5	-16342.6
kalign	-16212.4	-16218.3	-16208.2	-16206.9
mafft_einsi	-16178.6	-16185.1	-16179.3	-16178.6
mafft_ginsi	-16146.3	-16153.1	-16147.0	-16146.3
mafft_linsi	-16187.7	-16194.6	-16188.3	-16187.7
muscle	-16462.1	-16471.0	-16463.0	-16462.1
prank	-16069.9	-16078.5	-16070.3	-16069.5
probalign	-16447.7	-16454.6	-16448.0	-16446.7
probcons	-16373.4	-16377.8	-16374.1	-16374.0
promal_wPDB	-16382.0	-16382.0	-16378.5	-16377.6
model = mixed				
dialign	-16304.7	-16313.8	-16312.6	-16305.4
kalign	-16192.1	-16199.0	-16193.6	-16192.1
mafft_einsi	-16151.7	-16160.6	-16152.4	-16151.7
mafft_ginsi	-16113.0	-16120.3	-16113.6	-16113.0
mafft_linsi	-16154.2	-16160.5	-16154.9	-16154.2
muscle	-16425.4	-16433.6	-16427.0	-16425.1
prank	-16028.3	-16037.2	-16028.1	-16028.0
probalign	-16412.6	-16417.3	-16412.8	-16410.9
probcons	-16341.8	-16348.9	-16343.4	-16341.8
promal_wPDB	-16349.0	-16352.5	-16347.8	-16346.9

SI Fig 1. Previous hypotheses regarding phylogenetic relationships between gnathostome and cyclostome Hbs. According to the traditional view (A), the Hbs of cyclostomes and gnathostomes are orthologous proteins that derive from a proto-Hb precursor protein that had evolved an O_2-transport function in the vertebrate common ancestor (16). Under the 'parallel co-option or single co-option/secondary loss' hypothesis (B), the O_2-transport function evolved independently in both Hb lineages, or alternatively, an ancestral O_2-transport function was secondarily lost in the remaining gnathostome-specific globins (17). Finally, under the 'convergent co-option' hypothesis (C), O_2-transport functions evolved independently in the Hbs of cyclostomes and gnathostomes, or alternatively, an ancestral O_2-transport function was secondarily lost in gnathostome cytoglobin (18).



References

1 Ebner B, Burmester, T, Hankeln, T (2003) Globin genes are present in *Ciona intestinalis*. *Mol Biol Evol* 20:1521-1525.

2 Sullivan J, Joyce P (2005) Model Selection in Phylogenetics. *Annu Rev Ecol Evol Syst* 36:445-466.

3 Wong KM, Suchard MA, Huelsenbeck JP (2008) Alignment uncertainty and genomic analysis. *Science* 319:473-476.

4 Morgenstern B (2004) DIALIGN: multiple DNA and protein sequence alignment at BiBiServ. *Nucleic Acids Res* 32:33-36.

5 Lassmann T, Frings O, Sonnhammer ELL (2009) Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Res* 37:858-865.

6 Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9:286-298.

7 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797.

8 Löytynoja A, Goldman N (2005) An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci USA* 102:10557-10562.

9 Chikkagoudar S, Roshan U, Livesay D (2007) eProbalign: generation and manipulation of multiple sequence alignments using partition function posterior probabilities. *Nucleic Acids Res* 35:675-677.

10 Do CB, Mahabhashyam MSP, Brudno M, Batzoglou S (2005) ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res* 15:330-340.

11 Pei J, Tang M, Grishin NV (2008) PROMALS3D web server for accurate multiple protein sequence and structure alignments. *Nucleic Acids Res* 36:30-34.

12 Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8:275-282.

13 Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* 25:1307-1320.

14 Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4:18.

15 Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.

16 Goodman M, Pedwaydon J, Czelusniak J, Suzuki T, Gotoh T, Moens L, Shishikura F, Walz D, Vinogradov S (1988) An evolutionary tree for invertebrate globin sequences. *J Mol Evol* 27:236-249.

17 Burmester T, Hankeln T (2009) What is the function of neuroglobin? *J Exp Biol* 212:1423-1428.

18 Katoh K, Miyata T (2002) Cyclostome hemoglobins are possibly paralogous to gnathostome hemoglobins. *J Mol Evol* 55:246-249.